Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

WIP

Network architecture for Proton Decay Analyses in Liquid Argon Time Projection Chambers

KEVIN WIERMAN

PNNL
FNAL ML

Proton Decay in LArTPCs

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

1. Given a proposed process

arXiv:1512.06148v2

Decay Mode	Water Cherenkov		Liquid Argon TPC	
	Efficiency	Background	Efficiency	Background
$p \to K^+\bar{\nu}$	19%	4	97%	1
$p \to K^0\mu^+$	10%	8	47%	< 2
$p \to K^+\mu^-\pi^+$			97%	1
$n \to K^+e^-$	10%	3	96%	< 2
$n \to e^+\pi^-$	19%	2	44%	0.8

Numbers are based on studies done by A. Bueno et al. JHEP 04 (2007) 041

3. Get a measurement of n events

5. Goal is this limit

2. Simulate the background

$$P(\Gamma|n) = A\int_0^{\Gamma_{lim}} d\Gamma \int db \int db_{MC} \int d\lambda \int d\epsilon \frac{b_{MC}^{n_b}}{b} e^{-b_{MC}-(\Gamma\lambda\epsilon+b)-\frac{(b_{MC}-bC)^2}{2\sigma_b^2}} (\Gamma\lambda\epsilon + b)^n \ldots$$

4. Apply same cuts to data

Current issues with PDK

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

LineCluster — Identified as Particle

True Particle	μ	π	K	p
K	3.8%	5.0%	**44.8%**	3.7%
μ	**77.9%**	12.3%	0.4%	0.0%
e	**75.2%**	1.1%	0.4%	0.0%

TrajCluster — Identified as Particle

True Particle	μ	π	K	p
K	4.5%	4.2%	**47.9%**	7.6%
μ	**73.1%**	11.7%	0.4%	0.1%
e	**82.5%**	0.8%	0.2%	0.1%

Tingjun Yang, Dune Collaboration Meeting, January 2017

PIDA With CNN

- Utilize CNNs to Identify PDK events in data
  - $p \rightarrow K^+ \bar{\nu}$
  - $p \rightarrow \pi^+ \bar{\nu}$
  - $p \rightarrow e^+ \pi^0$
- Simulate these events and potential background processes

Backgrounds

- DUNE
  - Atmospheric Neutrinos
- MicroBooNE
  - Cosmics

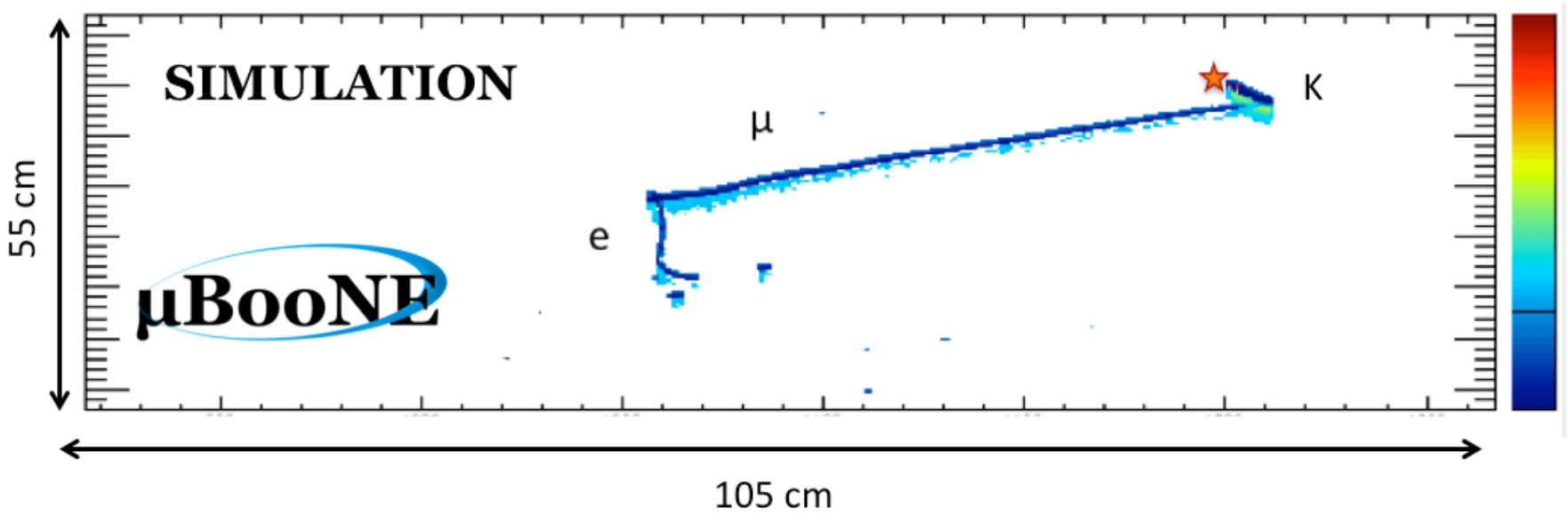
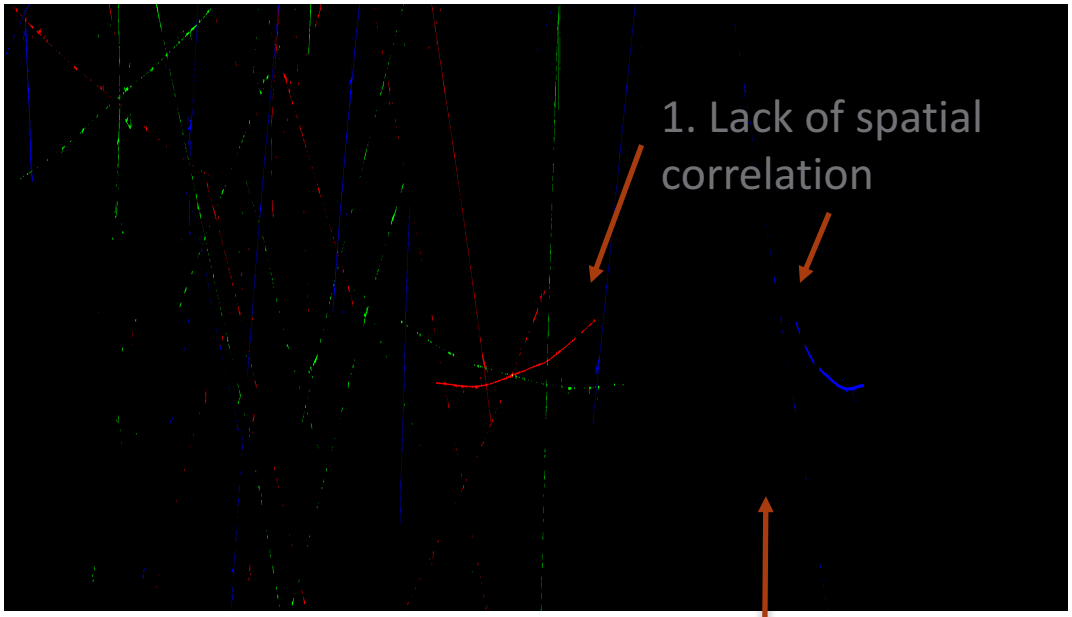


Image courtesy of Elena Gramellini.

Challenges in LArTPCs
Part 1: Data Preparation

- Need:
  - When feeding network
  - Tensor-like data
- Naïve approach:
  - R=U plane
  - G=V plane
  - B=Y/W/Z plane
  - Indices:
    - [filter,event,plane,wire,time]



2. This fraction of the image does Not show up in (u,v)

3. Size: Even for moderately sized TPCs, this is LARGE

Challenges in LArTPCs
Part 2: Categorization

- AKA forming the question
- Using MCParticle, the first order approach is:

For a label vector: $[0,1,3,0,6,0]$

With corresponding indices: $[e^-, e^+, \mu^-, \mu^+, K^+, K^-]$

This means: There are:
1 Positron
3 Muons
6 Kaons (+)

Following scheme developed by uBooNE collabroation
See Arxiv: https://arxiv.org/abs/1611.05531

Solution:
Ask a better question

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- The base question is:
  - Do we have a valid proton decay event?
    - First pass qualifications:
      - Are the final states correct?
      - Are the summed energies below mass shell?
- Now for second order qualification:
  - Need 3D tomography for summing total momenta to 0
    - For multi-particle final states
  - Need semantic segmentation for identifying which pixels to sum over

Solution:
Forming an Architecture from the data

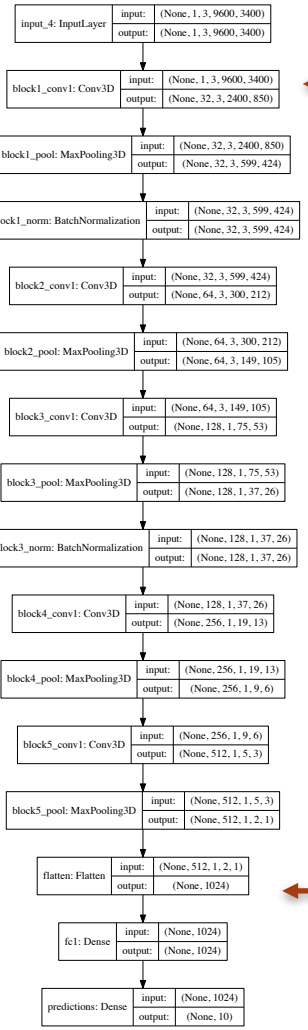
Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Separate Network into 2 parts:
  - 1. Intra-planar network
    - Should work uniformly on each layer, but not between layers
  - 2. Inter-planar network
    - Convolve on the planar axis

Creating a new CNN Architecture for LArTPCs

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Input



3x1 Convolution

Identify features common to all planes, but not ACROSS planes

Spatial Contraction

3x3 Convolution

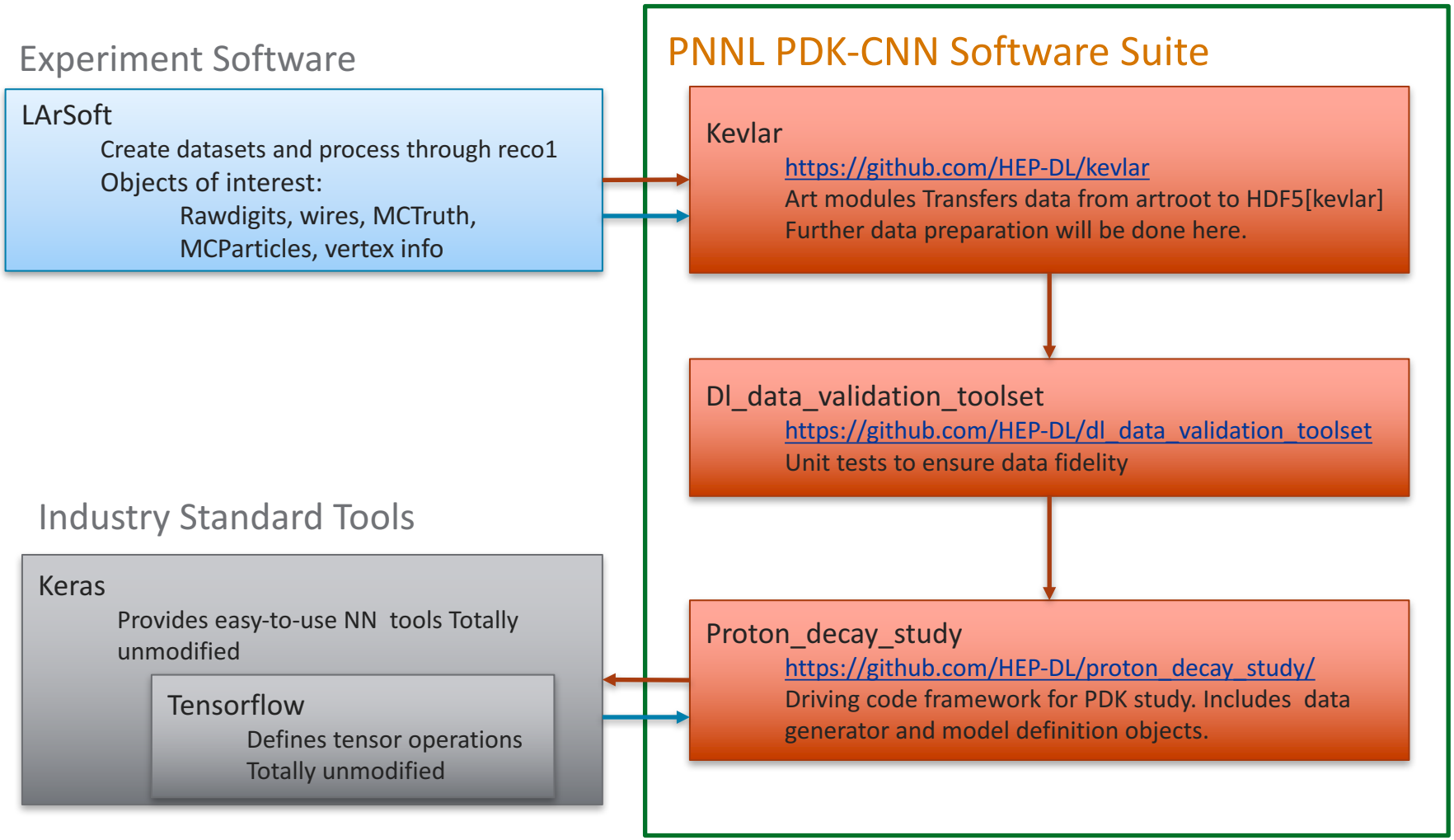
Identify features BETWEEN planes

Single Module:

- Conv3D
- MaxPooling
- Batch Renormalization

Flatten, FC and Softmax

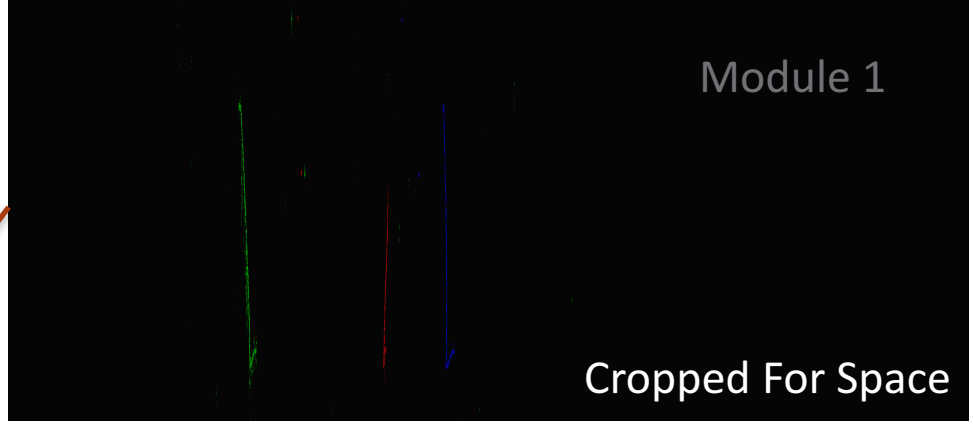
Underlying Toolset

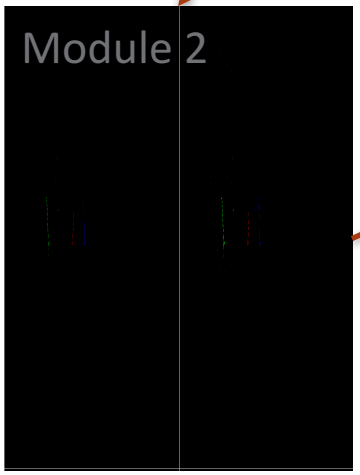


Training Scheme

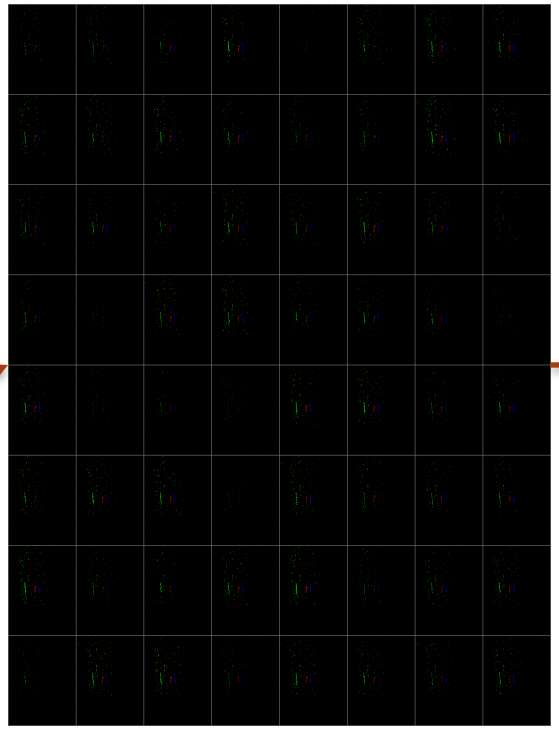
- Goal:
  - Create a common training framework for both uBooNE and DUNE
- In data production:
  - Create a LArSoft Dependent framework for generating MC samples
    - The Maze
      - https://github.com/kwierman/themaze
    - Does not depend on DUNE/uBooNE IP
    - Can create common samples for pretraining networks
- In training framework:
  - Create geometry-independent framework
  - Able to accept both uBooNE and DUNE geometry (currently single APA)

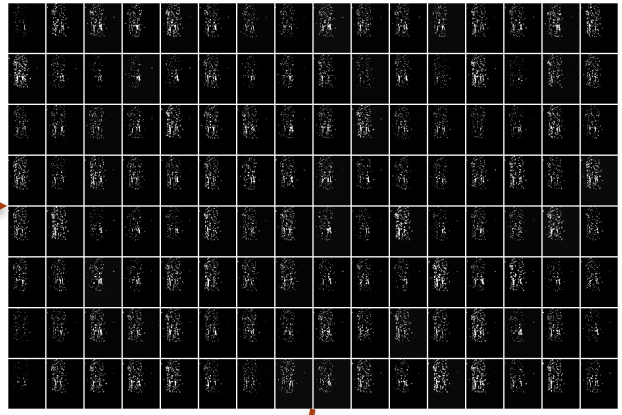
Feature Mapping

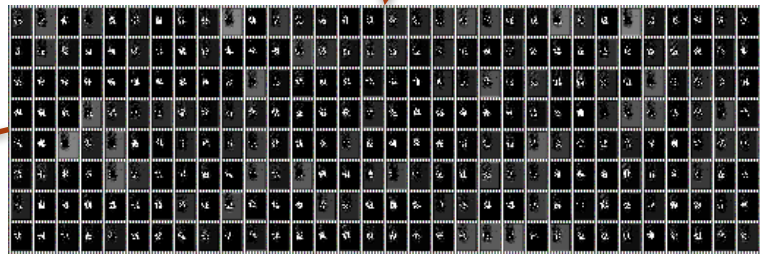




Also Cropped







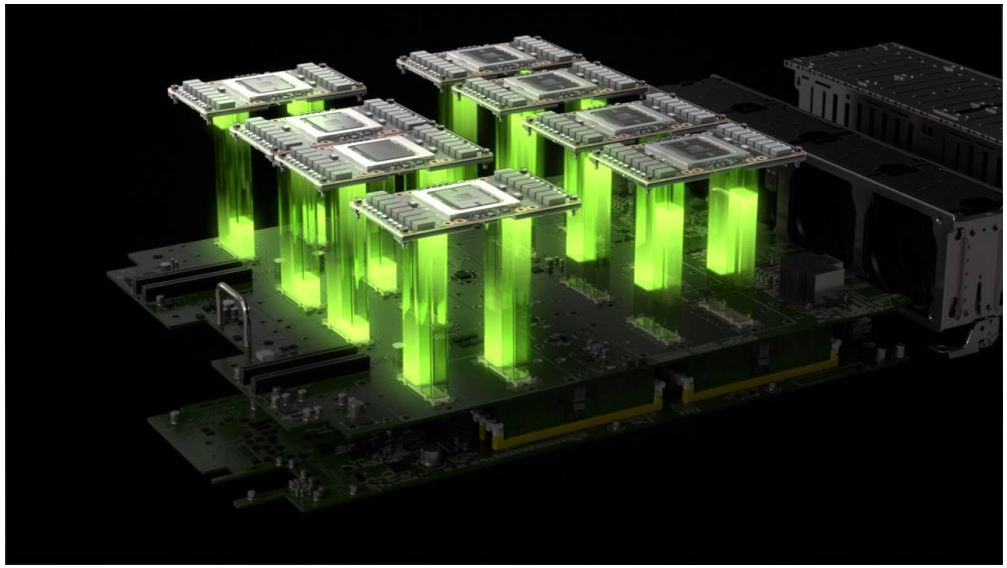
Current Status

1. For uBooNE:
    1. Prepare MC data sets
        1. Proton decay with new genie generator
        2. Cosmics with Corsika
        3. PDK with Corsika overlay
    2. Convert data to HDF5
        1. Create categorization
    3. Define model and data gen patterns
    4. Pretrain network on singles
    5. Train
        1. PDK (no cosmics)
        2. Cosmics (no pdk)
        3. Cosmics and PDK
2. For DUNE
    1. Prepare MC data sets
        1. Proton decay with new genie generator
            1. Geometry: 1x2x6
        2. Atmospheric neutrinos
    2. Convert data to HDF5
        1. No recob::wires yet

Work with Data Science and Analytics (DSA)

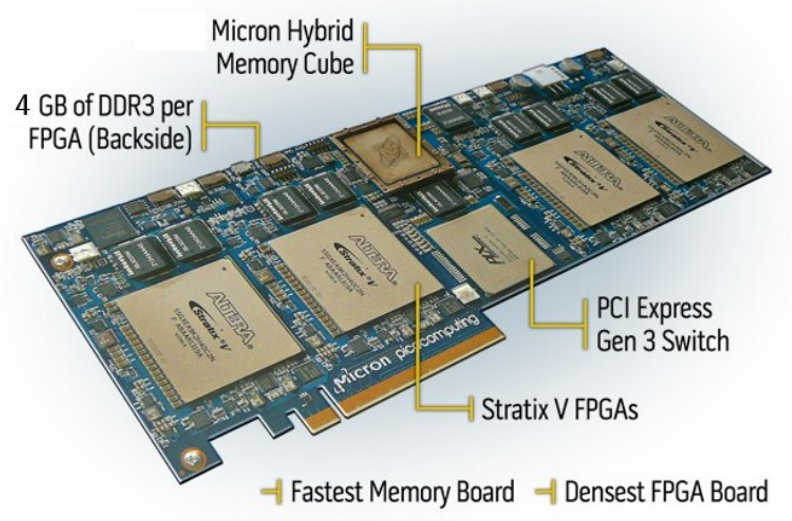
- Also working with DSA at PNNL
  - Access to several DGX-1's
    - Multi-GPU analysis
    - Necessary for full-sized images
  - Investigating a few other approaches
    - Sim-GAN
    - Wavelet decomposition
  - Developing Deep Learning techniques for multiple experiments
    - PNNL aims to be lead lab in DOE complex developing DL for mission science

DGX-1 Image courtesy of NVidia



DAQ

- PNNL developing scheme to identify events of interest
  - Using CNNs on FPGAs
    - For potential use as off beam-spill triggering
  - Have established an industry partner for developing this technology



Questions?

Data Preparation

For uBooNE:

- RGB maps to UVY
- Use computer vision

For DUNE:

- Really have to rethink how we deal with wrapped wires and collection plane
- The tensor operations don't change, but the tensor shapes do